

Transcriptional Analysis of the Developmental Stages of *Chlamydia trachomatis* A/HAR-13

Team H(oo)KD:
Hilda Delgadillo
Katrina Sherbina
Dillon Williams

11/12/2013

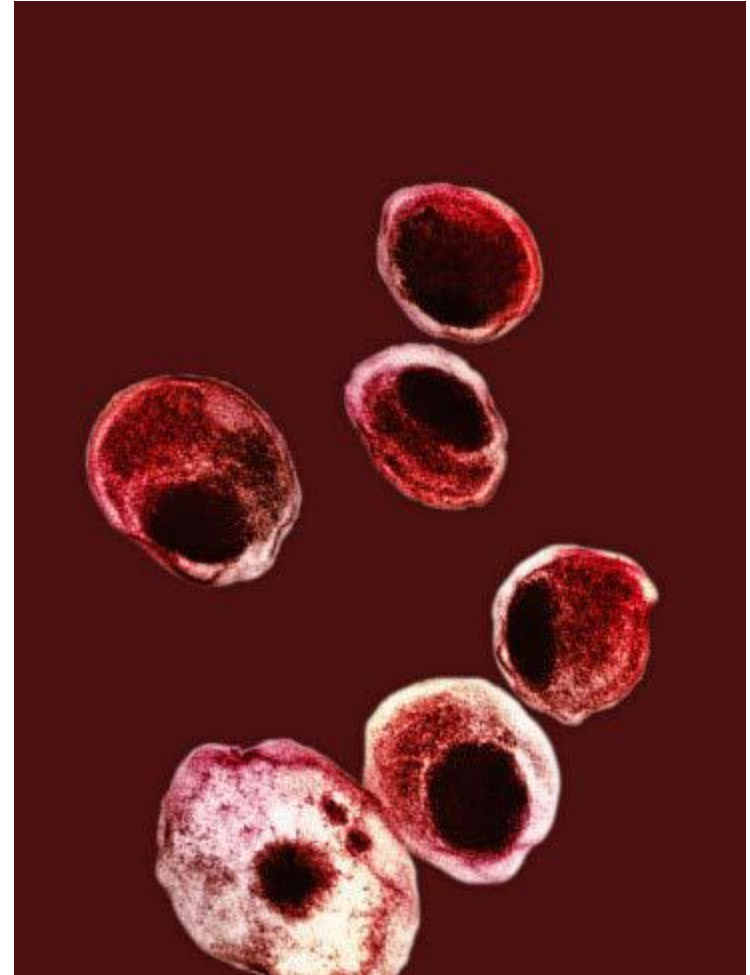
Loyola Marymount University

Outline

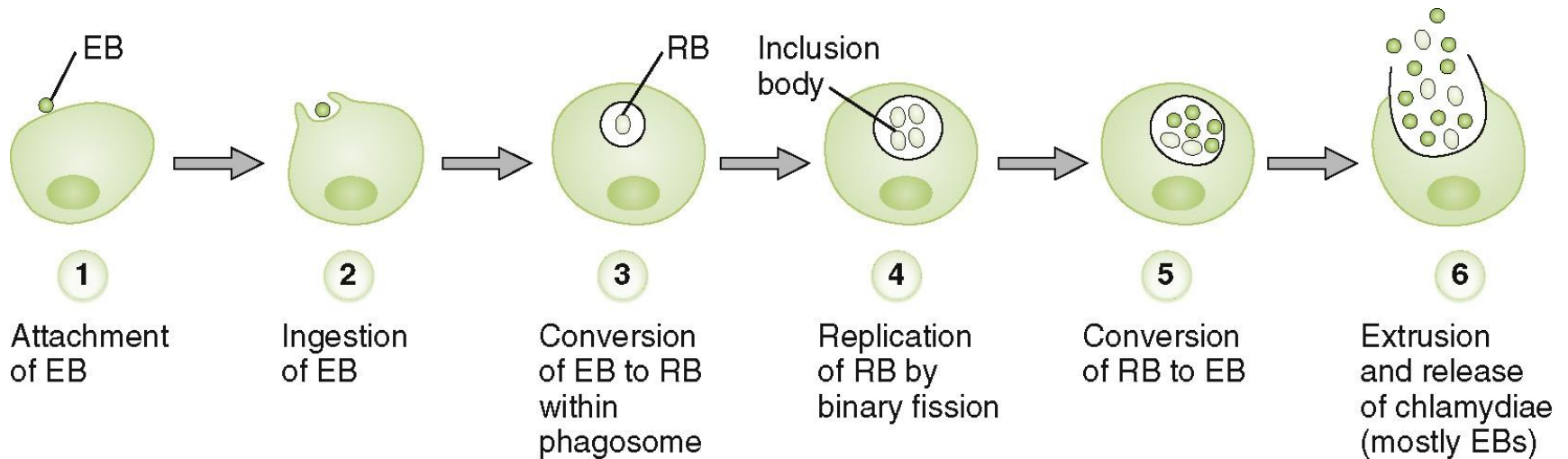
- *Chlamydia trachomatis* is an important human pathogen to study
- DNA microarray experiment sought to differentiate the developmental stages of *C. trachomatis*
- Gmbuilder 2.0b71 and pgAdminIII were used to create a gene database for *C. trachomatis*
- Database was vetted using TallyEngine, Access, SQL queries, XMLPipeDB match, and MOD
- Gmbuilder code was customized for *C. trachomatis*
- Raw microarray data was processed with Affymetrix software
- Statistical analysis of the microarray data was performed
- The data was analyzed with GenMAPP and MAPPFinder

Why is *Chlamydia trachomatis* important?

- Among the most clinically significant human pathogens
- Trachoma leading cause of infectious blindness-serious in developing countries
- Rise in number of infections in recent years
- Need for understanding of *Chlamydia trachomatis* virulence
- Obligate intracellular nature places restrictions



Developmental stages of *Chlamydia trachomatis*



<https://www.inkling.com/read/rapid-review-microbiology-and-immunology-rosenthal-tan-3rd/chapter-17/chlamydiae-and-zoonotic>

DNA microarray experiment sought to differentiate the developmental stages of *C. trachomatis*

- During infection of cultured host cells-nutrients (nucleotides, amino acids, and lipids)
- Host cell-free (axenic) medium- energy source requirements of EBs and RBs
- Rifampicin inhibits the conversion of developmental stage from EBs to RBs
- Serovar L2 - Sexually transmitted diseases
- Serovar L2 was analyzed with Serovar (A/HAR-13) DNA on the microarray chip
- Serovar A/HAR-13 - Blinding trachoma

Outline

- *Chlamydia trachomatis* is an important human pathogen to study
- DNA microarray experiment sought to differentiate the developmental stages of *C. trachomatis*
- Gmbuilder 2.0b71 and pgAdminIII were used to create a gene database for *C. trachomatis*
- Database was vetted using TallyEngine, Access, SQL queries, XMLPipeDB match, and MOD
- Gmbuilder code was customized for *C. trachomatis*
- Raw microarray data was processed with Affymetrix software
- Statistical analysis of the microarray data was performed
- The data was analyzed with GenMAPP and MAPPFinder

A gene database was built in pgAdminIII using gmbuilder 2.0b71

- Created database tables in pgAdminIII by running queries in *gmbuilder.sql*
- Imported UniProt XML, OBO-XML, and GOA files
- Exported a gene database
- Inspected the database
 - ▣ TallyEngine
 - ▣ XMLPipedDB Match
 - ▣ SQL query
 - ▣ OriginalRowCounts Table in Access

The TallyEngine count agreed with the SQL query count

TallyEngine

XML Path	XML Count	Database Table	Database Count
UniProt	917	UniProt	917
Ordered Locus	917	Ordered Locus	917
RefSeq	926	RefSeq	926
GeneID	926	GeneID	926
GO Terms	40071	GO Terms	40071

SQL Query

```
select count(*) from genenametype where type = 'ordered locus'  
and value ~ 'CTA_[0-9][0-9][0-9][0-9]';
```

Count: 917

Two types of gene IDs found in comparing the counts in Access and XMLPipeDB match

OriginalRowCounts Table in Access

Table	Rows
OrderedLocusNames	919
UniProt-Ordered Locus Names	919

XMLPipeDB Match

```
java -jar xmlpipedb-match-1.1.1.jar "CTA_[0-9][0-9][0-9][0-9]" <  
Uniprot_XML_C.trachomatis_serovar_A_KS_20131114.xml
```

Count: 911

```
java -jar xmlpipedb-match-1.1.1.jar "[p]CTA_[0-9][0-9][0-9][0-9]" <  
Uniprot_XML_C.trachomatis_serovar_A_KS_20131114.xml
```

Count: 9

Different counts resulted from the ordered locus ID CTA_0406/CTA_0407/CTA_0408

Counting Method	Count
TallyEngine	917
SQL Query	917
XMLPipeDB Match	919
OriginalRowCountsTable in Access	919
MOD	919

A species profile was created to customize the gmbuilder export for *C. trachomatis*

- Download GenMAPP Builder source code from SourceForge
- Code modified in Eclipse IDE for Java EE Developers + Subclipse
- Custom species profile code documents
 - ▣ Species name (*Genus species*)
 - ▣ Species specific database link:
http://bacteria.ensembl.org/chlamydia_trachomatis_a_har_13/Gene/Summary?g=~

The TallyEngine was customized for *C. trachomatis* within the gmbuilder code

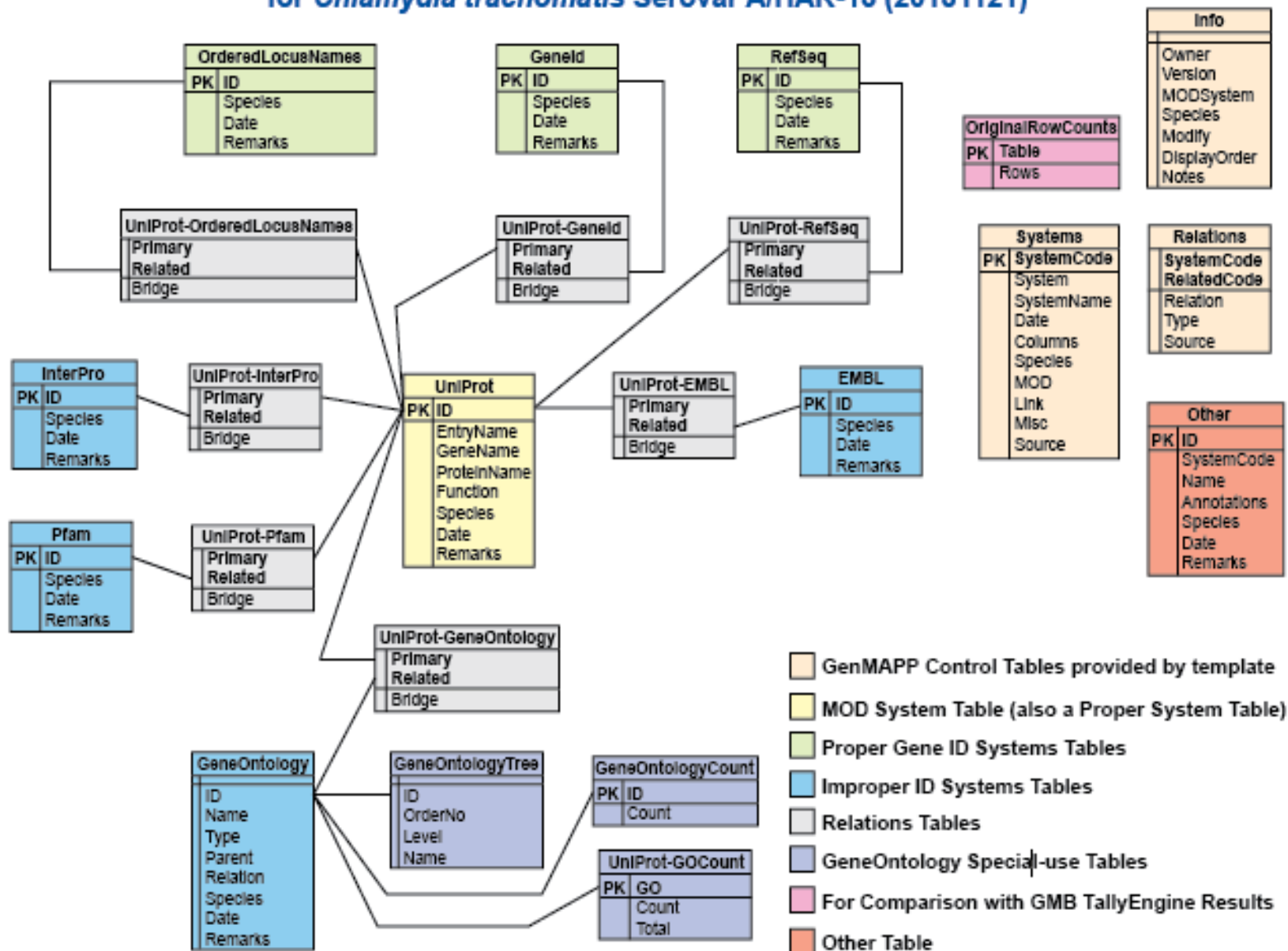
```
# Chlamydia trachomatis
chlamydiatrachomatisserovara_level_amount=1
chlamydiatrachomatisserovara_element_level0=uniprot/entry
    /gene/name&type&ordered locus
chlamydiatrachomatisserovara_query_level0=select count(*)
    from genenametype where type = 'ordered locus';
chlamydiatrachomatisserovara_table_name_level0=Ordered
    Locus
```

A gene database export was performed without receiving any errors

	UniProt XML	GO OBO-XML		GOA
File Name	uniprot-organism%3A315277+keyword%3A181.xml	go_daily-termdb.obo-xml.gz		22183.C_trachomatis_A.goa
Version #	UniProt release 2013_11	11/06/2013		11/12/13
Import Time	1.20 min	Import	13.05 min	0.04 min
		Processing	10.80 min	

.gdb File Name	Ct-Std v2 KS 20131121.gdb
Export Start Time	10:56 AM
Export End Time	11:14:26 AM

GenMAPP Gene Database Schema for *Chlamydia trachomatis* Serovar A/HAR-13 (20131121)



NOTE: Some Relations tables are not shown. All possible pairwise Relations tables exist between Proper ID systems and between Proper and Improper ID systems, but not between Improper ID systems (i.e., Proper-Proper, Proper-Improper, but NOT Improper-Improper).

The data needed to be modified to remove the Affymetrix IDs appended to each gene ID

	A	B	C	D
1	probe set	Axenic Media Rep 1	Axenic Media Rep 2	axenic media + rifampicin rep 2
2	AFFX-BioB-5_at	0.033567631	-0.255239492	0.229426198
3	AFFX-BioB-M_at	0.083116951	-0.148853078	-0.255944981
4	AFFX-BioB-3_at	0.223331716	-0.519291722	-0.288611893
5	AFFX-BioC-5_at	-0.069326758	-0.316781067	-0.394790036
6	AFFX-BioC-3_at	-0.111653351	-0.374826735	-0.443468371
11070	CpA_RQP02049_at	-0.451035698	-1.573735245	2.005167868
11071	CpA_RQP02052_at	1.383563587	0.191414424	-0.478957332
11072	CTA_0806_RRMH00303_at	-6.284612536	-4.151196656	-0.612940251
11073	CTA_0806_RRMH00303_at	-5.808779279	-5.22064971	-2.230443514
11074	CTA_0806_RRMH00303_at	-3.223003862	-5.244363835	-0.910732662
11075	CTA_0806_RRMH00303_at	-2.552405527	-2.553042199	-0.539108607
11076	CTA_0806_RRMH00303_at	-2.848495579	-2.338163065	-0.107650622
11077	CTA_0806_RRMH00303_at	-4.835519637	-5.142824185	-0.831567462
11078	CTA_0806_RRMH00303_at	-5.070854226	-4.022166285	-0.889817082
11079	CTA_0806_RRMH00303_at	-3.786810044	-3.397932754	0.373428723
11080	CTA_0806_RRMH00303_at	-2.491099815	-2.303624036	-0.732812693
11081	CTA_0806_RRMH00303_at	-4.064250191	-3.768191798	-0.884163214

Outline

- *Chlamydia trachomatis* is an important human pathogen to study
- DNA microarray experiment sought to differentiate the developmental stages of *C. trachomatis*
- Gmbuilder 2.0b71 and pgAdminIII were used to create a gene database for *C. trachomatis*
- Database was vetted using TallyEngine, Access, SQL queries, XMLPipeDB match, and MOD
- Gmbuilder code was customized for *C. trachomatis*
- Raw microarray data was processed with Affymetrix software
- Statistical analysis of the microarray data was performed
- The data was analyzed with GenMAPP and MAPPFinder

Affymetrix software was used to open raw microarray data

- Downloaded raw data from microarray website
 - ▣ Raw data was in the form of CEL files and used Affymetrix chips
 - ▣ Had to download dChip software in order to open raw data
 - ▣ Ran Log 2 transformation
 - ▣ Normalized arrays
 - ▣ Ran dChip software
 - ▣ Exported results into Excel spreadsheet

Sdrf file was used to identify experiment groups

- Raw Data did not have columns containing explicit identification regarding the data presented
- Found sdrf file (sample and data relationship format) on microarray website
 - ▣ Contained proper identification for gene ID's
 - ▣ Renamed columns based on gene ID information
 - E.g. GSM970952_4951_AO_8_Ct_01_EB-nt-co-rep1_01_5492 → Gene expression data from_EB grown in axenic media rep 1

Computed ratios of EB to RB in the presence and absence of rifampicin

- Following calculations were performed:
 - Avg. EB
 - Avg. RB
 - Avg. EB to RB
 - LogFC EB to RB
 - T statistic
 - P Value

There are more significant gene changes than would be expected by chance

Filter(s) Used	No Rifampicin	Rifampicin
p value < 0.05	864	471
p value < 0.01	794	308
p value < 0.001	676	93
p value < 0.0001	235	26
average log fold change > 0.25 & p value < 0.05	82	169
average log fold change < -0.25 & p value < 0.05	778	476

- Increase in a gene in EB due to inhibition of transcription from EB to RB
- Decrease in a gene in EB due to inhibition of transcription from RB back to EB

GenMAPP analysis was performed for the data collected in the absence of rifampicin

Gene Database

Ct-Std_v2_KS_20131121.gdb

Expression Dataset

Name: For_GenMAPP_Chlamydia_V4_20131205_KS

Color Set: EB to RB

Legend: EB to RB

- Increased
- Decreased
- No criteria met
- Not found

Expression Dataset Manager

Expression Datasets Color Sets Help

Name For_GenMAPP_Chlamydia_V4_20131205_KS

Remarks

Notes

Color Sets

Name EB to RB Gene Value LogFC_EB-to-RB

Criteria Builder

Columns: AverageEB, AverageRB, EB-to-RB, LogFC_EB-to-RB, P value

Ops: =, >, <, >=, <=, <>, AND, OR

Label in Legend: []

Color: []

Criterion: [LogFC_EB-to-RB] > 0.25 AND [P value] < 0.0

Label	Criterion	Color
1 Increased	[LogFC_EB-to-RB] > 0.25 AND [P value] < 0.0	Red
2 Decreased	[LogFC_EB-to-RB] < -0.25 AND [P value] < 0.0	Blue
3 No criteria met		Grey
4 Not found		White

Buttons: New, Save, Add, Move Up, Move Down, Edit, Delete

MAPPFinder results for the increased criterion

GO Terms Significantly Increased	Z Score	P Value	Number Changed	Percent Changed
pathogenesis	3.59	0.017	3	27.27273
multi-organism process	3.59	0.017	3	27.27273
macromolecule modification	2.707	0.018	5	13.88889

Z score is greater than 2.

P Value is less than 0.05.

Number changed filter is greater than or equal to 3 & less than 150.

Percent changed is greater than 13.

MAPPFinder results for the decreased criterion

GO Terms Significantly Decreased	Z Score	P Value	Number Changed	Percent Changed
macromolecule biosynthetic process	2.967	0.002	146	96.68874
cellular macromolecule biosynthetic process	2.944	0.002	145	96.66666
translation	2.902	0.004	88	98.8764
gene expression	2.705	0.004	135	96.42857
metal ion binding	2.76	0.013	82	98.79518
cation binding	2.435	0.021	84	97.67442
cytoplasmic part	2.004	0.034	68	97.14286
lipid metabolic process	2.027	0.035	37	100
cellular lipid metabolic process	2.027	0.035	37	100
carbohydrate derivative metabolic process	2.252	0.04	77	97.46835
organonitrogen compound metabolic process	2.098	0.043	123	95.34884
macromolecular complex	2.039	0.043	83	96.51163
organophosphate biosynthetic process	2.169	0.045	42	100

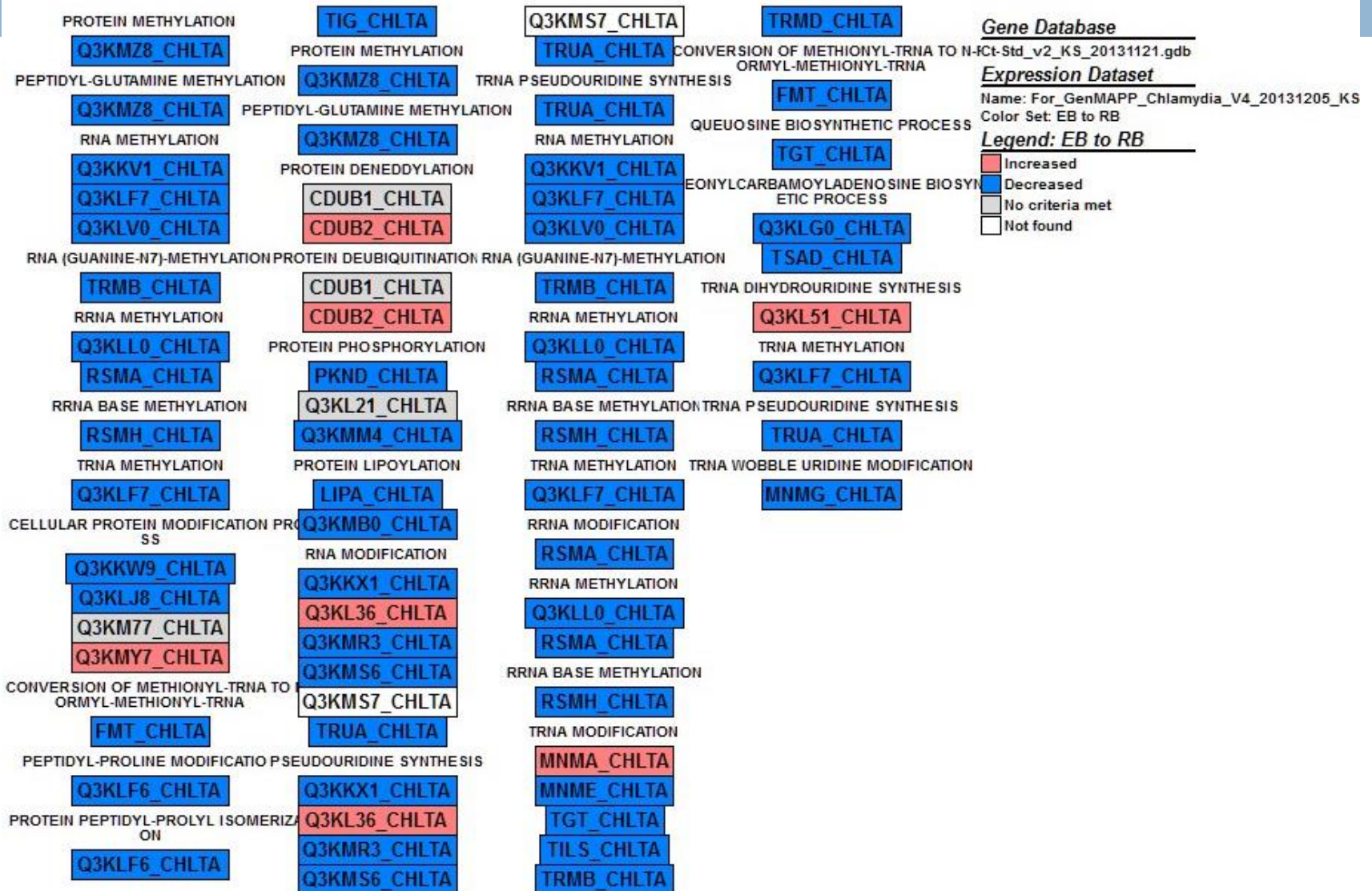
Z score is greater than 2.

P Value is less than 0.05.

Number changed filter is greater than or equal to 3 & less than 150.

Percent changed is greater than 50.

Macromolecule modification MAPP demonstrates that EBs are not metabolically dormant



Conclusion

- The gmbuilder process for *C. trachomatis* had few problems
- Differences in the counts were due to the separation or lack thereof of the ID CTA_0406/CTA_0407/CTA_0408
- Minimal modifications were made to the gmbuilder code
- There are more significant gene changes than would be expected by chance
- Rifampicin has a notable effect on gene expression
- MAPPFinder analysis shows that EBs are not metabolically dormant